



## Bilag – De socioøkonomiske referencer for gymnasiekarakterer 2021

### Bilag 1: Socioøkonomiske baggrundsoplysninger

Bilagstabel 1 Baggrundsoplysninger

Baggrundsoplysning	Valgt kodning
<b>9. klasse FSA karaktergennemsnit</b>	Gennemsnit af bundne prøver (ekskl. dansk or- den) <sup>1</sup> .
<b>Køn</b>	Dreng Pige
<b>Alder på tidspunkt for afslutning af uddannelsen</b>	18 år eller yngre 19 år 20 år 21 år eller ældre
<b>Herkomst<sup>2</sup></b>	Dansk Indvandrer Efterkommer

<sup>1</sup> Eleven skal have karakter i mindst 4 ud af 8 (2007->) eller 4 ud af 7 (2002-06) bundne prøvefags karakterer.

<sup>2</sup> Der skelnes mellem indvandrere, efterkommere, og personer med dansk oprindelse. Personer med dansk oprindelse (dansk) er personer, hvor mindst en af forældrene er dansk statsborger og født i Danmark.

Efterkommere er født i Danmark. Ingen af forældrene er danske statsborgere, født i Danmark. Når en eller begge forældre, der er født i Danmark, opnår dansk statsborgerskab, vil deres børn ikke blive klassificeret som efterkommere. Fastholder danskfødte forældre imidlertid begge et udenlandsk statsborgerskab, vil deres børn blive klassificeret som efterkommere. Hvis der ikke findes oplysninger om nogen af forældrene, og personen er udenlandsk statsborger, opfattes personen også som efterkommer.

Indvandrere er født i udlandet. Ingen af forældrene er danske statsborgere, født i Danmark. Hvis der ikke findes oplysninger om nogen af forældrene, og personen er født i udlandet, opfattes personen som indvandrer.

<b>Forældrenes højeste fuldførte uddannelse<sup>3</sup></b>	Grundskolen Gymnasial uddannelse Erhvervsfaglig uddannelse Kort videregående uddannelse Mellemlang videregående uddannelse Lang videregående uddannelse (inkl. ph.d. mv.)
---	--

**Bilagstabel 1 Baggrundsoplysninger (fortsat)**

Baggrundsoplysning	Valgt kodning
<b>Forældrenes gennemsnitlige bruttoindkomst<sup>4</sup></b>	- 300.000 kr. 300.000 - 400.000 kr. 400.000 - 500.000 kr. 500.000+ kr.
<b>Familiestatus<sup>5</sup></b>	Hjemmeboende hos par Hjemmeboende hos enlig Udeboende
<b>Adgangsvej<sup>6</sup></b>	Folkeskole Friskole og privat grundskole Efterskole Anden
<b>10. klasse<sup>7</sup></b>	Nej

<sup>3</sup> Den højeste fuldførte uddannelse blandt elevens mor og far oplyst ved start på uddannelsen. Hvis uoplyst da indgår oplysningen fra det år eleven fyldte 13 år.

<sup>4</sup> Bruttoindkomsten er i året for start på uddannelsen.

<sup>5</sup> Familiestatus i året for start på uddannelsen er en kombination af familietype og plads. Familietypen par indeholder ægtepar, registreret partnerskab, samlevende og samboende par. Plads er enten hjemmeboende eller udeboende. Hvis uoplyst da indgår oplysningen fra det år eleven fyldte 13 år.

<sup>6</sup> Sidste grundskoletype (9. klasse eller afsluttet 10. klasse) inden start på uddannelsen.

<sup>7</sup> Gennemført 10. klasse.

Ja

---

## Bilag 2: Den statistiske model

Den anvendte statistiske model er en såkaldt multilevel model:

$$y_{ij} = x_{ij}\beta + u_j + e_{ij}$$

hvor  $y_{ij}$  er karakteren for elev  $i$  på institution  $j$ ,  $x_{ij}$  er elevens baggrundsvariable og  $\beta$  er de tilhørende parameterestimater,  $u_j$  er varianskomponenten svarende til variationen mellem institutionerne, og  $e_{ij}$  er residualen svarende til variationen mellem eleverne på institutionen. For hver kombination af fag, prøveform og niveau foregår beregningerne på individniveau for hver af de gymnasiale uddannelser hver for sig.

Ved anvendelse af modellen fås et estimat over, hvor stor en del af variationen mellem karaktererne der skyldes forskelle mellem eleverne, og hvor stor en del der skyldes forskelle mellem institutionerne. Dette kaldes modellens forklaringsgrad.

Alle de socioøkonomiske baggrundsvariable beskrevet i bilag 1 (Bilagstabel 1) medtages i modellen sammen med interaktionen mellem forældrenes uddannelse og elevens herkomst. Desuden medtages institutionsniveau variable i form af andelen af indvandrere og efterkommere på institutionen, andelen af elever med forældre hvis højeste fuldførte uddannelse mindst er en mellemlang videregående uddannelse samt andelen af elever med forældre, der har en lav gennemsnitlig bruttoindkomst. Interaktionen medtages for, at undersøge om effekten af forældrenes uddannelse afhænger af elevens herkomst. Institutionsniveau variablene medtages for at undersøge, hvorvidt der er en 'klassekammerat effekt', dvs. om der er en øget effekt af, at der er en høj/lav andel af elever med en bestemt baggrund. Øvrige interaktioner og institutionsniveau variable er ikke medtaget dels for ikke at øge modellernes kompleksitet, dels viser undersøgelser, at disse ikke øger modellernes forklaringskraft væsentligt. I den statistiske model for beregning af socioøkonomiske referencer for 3-års perioden medtages endvidere dimittendåret som en forklarende variabel. Herved tages der højde for, at karakterniveauet kan være forskelligt fra år til år.

Følgende baggrundsvariable er ikke statistisk signifikante i den multiple model og medtages derfor ikke i den endelige beregning af de socioøkonomiske referencer: forældrenes arbejdsmarkedsstatus samt alle institutionsniveau variablene. Når der således er taget højde for elevernes individuelle baggrundsplysninger, kan der ikke i disse modeller konstateres en øget effekt af, at der er en høj/lav andel af elever med en bestemt baggrund på den enkelte institution.

Ikke alle elever har fyldestgørende oplysninger på de her benyttede baggrundsvariable. En ofte anvendt metode til håndtering af dette er 'missing value' indikator metoden, hvor uoplyst medtages som en selvstændig kategori på lige fod med baggrundsvariablenes øvrige kategorier, og der så tilføjes en indikatorvariabel, der fortæller hvorvidt, der er tale om en uoplyst værdi.

Omfanget af ukendt oplysning er dog ikke uafhængig af herkomst. Oplysningsgraden for alle baggrundsvariablene er størst for elever af dansk herkomst. Der er således ikke tale om tilfældige 'huller' i informationen. Det er fx specielt udtalt blandt

indvandrere, at informationen om forældrenes uddannelse mangler. Da uddannelsesniveaueet blandt de indvandrere og efterkommere, hvor uddannelsesoplysningerne kendes, er en del anderledes end blandt danskere, vil der være en stor risiko for, at 'missing value' indikator metoden fejlesterimerer effekten af såvel herkomst som af uddannelse. Tilsvarende problematik eksisterer for de øvrige baggrundsvariable.

Derfor er der i stedet valgt at indsætte en værdi, et såkaldt bedste gæt, på alle de steder, hvor oplysningerne er ukendte. For hver elev med ukendt baggrundsoplysning gættes 5 gange, således at modellen bliver estimeret på baggrund af en sandsynlig fordeling. Metoden kaldes 'multiple imputation'. Gættene baseres på fordelingen blandt elever med oplyst baggrundsvariabel, hvor der er taget hensyn til elevens køn, alder og herkomst. Den socioøkonomiske reference for hver elev er så et gennemsnit af de 5 beregnede værdier.

For hver elev beregnes en socioøkonomisk reference ud fra værdierne af elevens baggrundsvariable og de estimerede parametre for baggrundsvariablene. Herefter beregnes forskellen mellem opnået karakter og den socioøkonomiske reference, det såkaldte residual  $r_{ij}$ , for hver elev.

På hovedinstitutionsniveau er den estimerede forskel mellem karaktergennemsnittet og den socioøkonomiske reference i en ordinær regressionsmodel gennemsnittet af disse elevresidualer på den enkelte institution,  $\bar{r}_j$ . I de her anvendte to-niveau modeller, hvor vi både har variation mellem institutioner ('Between'), og variation mellem elever indenfor institutionen ('Within'), bliver residualerne på hovedinstitutionsniveau:

$$\hat{u}_j = c_j \bar{r}_j$$

hvor  $c_j$  er den såkaldte 'shrinkage' faktor, defineret ved

$$c_j = \frac{B}{(B + W/n_j)}$$

$B$  er variationen mellem institutioner ('Between'),  $W$  er residualvariationen mellem elever indenfor institutionen ('Within') og  $n_j$  er antallet af elever på den  $j$ 'te institution.

For forskellen mellem opnået karakter og den socioøkonomiske reference for hver hovedinstitution er der endvidere udregnet et 95 % sikkerhedsinterval. Hvis dette sikkerhedsinterval ligger over 0, opnår institutionens elever et statistisk signifikant højere karaktergennemsnit end elever på landsplan med lignende baggrundsforhold, mens det modsatte er tilfældet, hvis sikkerhedsintervallet ligger under 0. Indeholder 95 % sikkerhedsintervallet værdien 0, da opnår eleverne på den pågældende institution et karaktergennemsnit som elever på landsplan med lignende baggrundsforhold.

Sikkerhedsintervallet afhænger dels af den estimerede variation og dels af antallet af elever, der indgår i beregningerne for den enkelte institution. Således bliver sikkerhedsintervallerne mindre for den enkelte institution, når der betragtes en 3-årig periode, idet der her indgår 3 dimittendårgange i beregningerne.

I bilag 3 fremgår estimerne for de socioøkonomiske baggrundsvariable for eksamensresultatet inkl. bonus A. De viste estimer vedrører de simple modeller, hvor kun én baggrundsvariabel er medtaget af gangen. Estimerne i disse modeller giver et indtryk af størrelsesordenen af 'effekten' af den enkelte baggrundsvariabel. I den endelige model, hvor alle baggrundsvariablene er medtaget, er effekterne ikke uafhængig af hinanden og kan derfor vanskeligt fortolkes enkeltvis.

Ved anvendelse af den statistiske model fås endvidere et estimat over, hvor stor en del af variationen mellem karaktererne der skyldes forskelle mellem eleverne, og hvor stor en del der skyldes forskelle mellem institutionerne. Den del af variationen, der kan forklares af baggrundsvariablene, kaldes modellens forklaringsgrad. I bilag 3 fremgår også modellens forklaringsgrad i disse simple modeller. Her ses hvor stor del af variationen i elevernes karakterer som kan forklares af den enkelte socioøkonomiske baggrundsvariabel.

## Bilag 3: Parameterestimer og forklaringsgrad

I bilagstabel 2 - bilagstabel 5 findes parameterestimerne for de enkelte socioøkonomiske baggrundsvariable i modellerne for eksamensresultatet inkl. bonus A for henholdsvis hf, hhx, htx og stx. De viste estimer vedrører de såkaldte simple modeller, hvor kun én baggrundsvariabel er medtaget af gangen.

Estimerne i de endelige (multiple) modeller med samtlige baggrundsvariable kan afvige fra de viste, da de socioøkonomiske baggrundsvariable ikke er uafhængige. Fx er forældrenes højeste fuldførte uddannelse og forældrenes bruttoindkomst ikke uafhængige. Det samme gælder karaktergennemsnittet fra 9. klasse og de øvrige baggrundsvariable.

Endvidere er angivet den del af variationen i elevernes karakterer, der kan forklares af den pågældende socioøkonomiske baggrundsvariabel.

Bilagstabel 2 Hf 2021. Parameterestimer (standard error) og forklaringsgraden for de enkelte socioøkonomiske baggrundsvariable. Eksamensresultat (inkl. bonus A).

Baggrundsplysning	Kodning	Estimat (SE)	Forklaringsgrad (%)
<b>Karaktergennemsnit 9. klasse FSA</b>		0,76 (0,01)	38 %
<b>Køn</b>	Dreng	0	1 %
	Pige	0,59 (0,06)	
<b>Alder</b>	18 år eller yngre	0	1 %
	19 år	0,03 (0,07)	
	20 år	0,14 (0,09)	
	21 år eller ældre	0,54 (0,09)	
<b>Herkomst</b>	Dansk	0	7 %
	Indvandrer	-1,27 (0,14)	
	Efterkommer	-1,67 (0,09)	
<b>Forældrenes højeste fuldførte uddannelse</b>	Grundskolen	0	6 %
	Gymnasial uddannelse	0,30 (0,17)	
	Erhvervsfaglig uddannelse	0,47 (0,10)	
	Kort videregående uddannelse	0,89 (0,13)	
	Mellemlang videregående uddannelse	1,29 (0,11)	
<b>Forældrenes gennemsnitlige bruttoindkomst</b>	Lang videregående uddannelse	1,79 (0,13)	5 %
	- 300.000 kr.	0	
	300 - 400.000 kr.	0,65 (0,08)	
	400 - 500.000 kr.	1,02 (0,08)	
<b>Familiestatus</b>	500.000+ kr.	1,26 (0,08)	< 1 %
	Hjemmeboende hos par	0	
	Hjemmeboende hos enlig	-0,00 (0,07)	
<b>Adgangsvej</b>	Udeboende	0,28 (0,08)	2 %
	Folkeskolen	0	
	Friskole og privat grundskole	0,49 (0,10)	
	Efterskole	0,72 (0,07)	
<b>10. klasse</b>	Andet	0,07 (0,12)	1 %
	Nej	0	
	Ja	-0,34 (0,06)	



**Bilagstabel 3 Hhx 2021. Parameterestimer (standard error) og forklaringsgraden for de enkelte socioøkonomiske baggrundsvariable. Eksamensresultat (inkl. bonus A).**

Baggrundsplysning	Kodning	Estimat (SE)	Forklaringsgrad (%)
Karaktergennemsnit 9. klasse FSA		0,87 (0,01)	46 %
Køn	Dreng	0	6 %
	Pige	1,20 (0,05)	
Alder	18 år eller yngre	0	2 %
	19 år	-0,42 (0,07)	
	20 år	-0,84 (0,08)	
	21 år eller ældre	-2,17 (0,18)	
Herkomst	Dansk	0	2 %
	Indvandrer	-1,10 (0,21)	
	Efterkommer	-1,29 (0,11)	
Forældrenes højeste fuldførte uddannelse	Grundskolen	0	4 %
	Gymnasial uddannelse	0,70 (0,17)	
	Erhvervsfaglig uddannelse	0,83 (0,14)	
	Kort videregående uddannelse	1,38 (0,16)	
	Mellemlang videregående uddannelse	1,25 (0,15)	
Forældrenes gennemsnitlige bruttoindkomst	Lang videregående uddannelse	2,09 (0,16)	4 %
	- 300.000 kr.	0	
	300 - 400.000 kr.	0,29 (0,09)	
	400 - 500.000 kr.	0,59 (0,09)	
Familiestatus	500.000+ kr.	1,18 (0,08)	1 %
	Hjemmeboende hos par	0	
	Hjemmeboende hos enlig	-0,42 (0,06)	
Adgangsvej	Udeboende	-1,01 (0,17)	1 %
	Folkeskolen	0	
	Friskole og privat grundskole	0,29 (0,08)	
	Efterskole	0,32 (0,06)	
10. klasse	Andet	-1,10 (0,20)	1 %
	Nej	0	
	Ja	-0,40 (0,05)	

**Bilagstabel 4 Htx 2021. Parameterestimer (standard error) og forklaringsgraden for de enkelte socioøkonomiske baggrundsvariable. Eksamensresultat (inkl. bonus A).**

Baggrundoplysning	Kodning	Estimat (SE)	Forklaringsgrad (%)
Karaktergennemsnit 9. klasse FSA		0,83 (0,01)	45 %
Køn	Dreng	0	4 %
	Pige	1,01 (0,08)	
Alder	18 år eller yngre	0	2 %
	19 år	-0,12 (0,09)	
	20 år	-0,65 (0,10)	
	21 år eller ældre	-1,67 (0,20)	
Herkomst	Dansk	0	1 %
	Indvandrer	-1,01 (0,20)	
	Efterkommer	-0,72 (0,13)	
Forældrenes højeste fuldførte uddannelse	Grundskolen	0	5 %
	Gymnasial uddannelse	0,63 (0,28)	
	Erhvervsfaglig uddannelse	0,72 (0,19)	
	Kort videregående uddannelse	1,14 (0,21)	
	Mellemlang videregående uddannelse	1,31 (0,18)	
Forældrenes gennemsnitlige bruttoindkomst	Lang videregående uddannelse	2,06 (0,19)	3 %
	- 300.000 kr.	0	
	300 - 400.000 kr.	0,50 (0,12)	
	400 - 500.000 kr.	0,61 (0,11)	
Familiestatus	500.000+ kr.	1,17 (0,11)	1 %
	Hjemmeboende hos par	0	
	Hjemmeboende hos enlig	-0,23 (0,09)	
Adgangsvej	Udeboende	-1,10 (0,22)	1 %
	Folkeskolen	0	
	Friskole og privat grundskole	0,18 (0,10)	
	Efterskole	0,32 (0,08)	
10. klasse	Andet	-1,12 (0,23)	< 1 %
	Nej	0	
	Ja	-0,34 (0,07)	

Bilagstabel 5 Stx 2021. Parameterestimer (standard error) og forklaringsgraden for de enkelte socioøkonomiske baggrundsvariable. Eksamensresultat (inkl. bonus A).

Baggrundsplysning	Kodning	Estimat (SE)	Forklaringsgrad (%)
<b>Karaktergennemsnit 9. klasse FSA</b>		0,93 (0,01)	56 %
<b>Køn</b>	Dreng	0	2 %
	Pige	0,78 (0,03)	
<b>Alder</b>	18 år eller yngre	0	1 %
	19 år	-0,15 (0,03)	
	20 år	-0,44 (0,04)	
	21 år eller ældre	-1,45 (0,15)	
<b>Herkomst</b>	Dansk	0	3 %
	Indvandrer	-0,85 (0,11)	
	Efterkommer	-1,22 (0,05)	
<b>Forældrenes højeste fuldførte uddannelse</b>	Grundskolen	0	12 %
	Gymnasial uddannelse	0,85 (0,12)	
	Erhvervsfaglig uddannelse	0,88 (0,08)	
	Kort videregående uddannelse	1,33 (0,09)	
	Mellemlang videregående uddannelse	1,73 (0,08)	
<b>Forældrenes gennemsnitlige bruttoindkomst</b>	- 300.000 kr.	0	8 %
	300 - 400.000 kr.	0,36 (0,05)	
	400 - 500.000 kr.	0,88 (0,05)	
	500.000+ kr.	1,56 (0,05)	
<b>Familiestatus</b>	Hjemmeboende hos par	0	1 %
	Hjemmeboende hos enlig	-0,48 (0,04)	
	Udeboende	-0,64 (0,12)	
<b>Adgangsvej</b>	Folkeskolen	0	2 %
	Friskole og privat grundskole	0,31 (0,04)	
	Efterskole	0,57 (0,03)	
	Andet	-1,17 (0,14)	
<b>10. klasse</b>	Nej	0	< 1 %
	Ja	-0,12 (0,03)	